

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-207170

(43)Date of publication of application : 28.07.2000

(51)Int.Cl.

G06F 3/16

G10L 11/00

(21)Application number : 11-008195

(71)Applicant : SONY CORP

(22)Date of filing : 14.01.1999

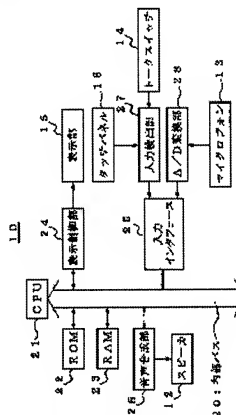
(72)Inventor : SASAI TAKASHI
HATTORI MASAKAZU
TSUNODA HIROSHI
KATO YASUHIKO

(54) DEVICE AND METHOD FOR PROCESSING INFORMATION

(57)Abstract:

PROBLEM TO BE SOLVED: To present information easy to comprehend by users and to exchange additional information or high-accuracy information with a few errors between devices.

SOLUTION: A CPU 21 generates an output audio signal by adding the additional information to an audio signal so as not to affect the hearing of voices due to that audio signal. The additional information is information related with the audio signal, for example. The voice composed of this output audio signal is outputted from a loudspeaker 12 towards an opposite side device. The opposite side device obtains an input audio signal by fetching that voice from a microphone 13 and supplies this audio signal to the CPU 21. The CPU 21 extracts the additional information from that input audio signal and performs processing, based on that additional information. For example, a phrase included in a natural language expressed by the voice is displayed on a display part 15. Furthermore, for example, voice recognizing processing is performed while limiting the object area of the natural language expressed by the voice with the additional information.



*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.*** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1]An information processor comprising:

An audio signal generating means which generates an audio signal.

An additional information generating means which generates additional information.

An information adding means which adds the above-mentioned additional information to the above-mentioned audio signal in a mode which does not influence listening comprehension of a sound by an audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne., and generates an output sound signal.

A voice output means which outputs a sound by the above-mentioned output sound signal.

[Claim 2]The information processor according to claim 1, wherein the above-mentioned additional information is information relevant to the above-mentioned audio signal.

[Claim 3]The information processor according to claim 2 which the above-mentioned audio signal expresses predetermined natural language, and is characterized by the above-mentioned additional information being information which shows words and phrases contained in the above-mentioned natural language.

[Claim 4]The information processor according to claim 2 which the above-mentioned audio signal expresses predetermined natural language, and is characterized by the above-mentioned additional information being information used for recognition or an interpretation of the above-mentioned natural language.

[Claim 5]The information processor according to claim 4, wherein the above-mentioned additional information is information which shows a keyword relevant to the above-mentioned natural language.

[Claim 6]The information processor according to claim 4, wherein the above-mentioned additional information is information which shows an intention or feeling expressed by the above-mentioned natural language.

[Claim 7]The information processor according to claim 4, wherein the above-mentioned additional information is information for identifying an object domain where the above-mentioned natural language is included.

[Claim 8]The information processor according to claim 2 which the above-mentioned audio signal expresses a place with natural language, and is characterized by the above-mentioned additional information being position information which shows the above-mentioned place.

[Claim 9]The information processor according to claim 2 which the above-mentioned audio signal expresses predetermined natural language, and is characterized by the above-mentioned additional information being information on foreign country language corresponding to the above-mentioned natural language.

[Claim 10]The information processor according to claim 2, wherein the above-mentioned audio signal is information which shows intention or feeling that the above-mentioned additional information is expressed by the above-mentioned voice corresponding to voice of human being or an animal.

[Claim 11]The information processor according to claim 2, wherein the above-mentioned audio signal is information the above-mentioned additional information indicates a kind and a name of the above-mentioned animal to be corresponding to a cry of an animal.

[Claim 12]The information processor according to claim 1, wherein the above-mentioned additional information is identification information which shows a self device.

[Claim 13]The information processor according to claim 1, wherein the above-mentioned additional information is information for identifying a device which has the right to output the above-mentioned sound next.

[Claim 14]The information processor comprising according to claim 1:

An attack primary detecting element of an audio signal where the above-mentioned information adding means is rapid and which stands up and detects a large amplitude portion as an attack. A spectrum-analysis part which conducts a spectrum analysis about the section of length when the above-mentioned audio signal was defined beforehand.

A hits section formation part which the above-mentioned attack portion in the above-mentioned audio signal is removed from an output of the above-mentioned attack primary detecting element, and an output of the above-mentioned spectrum-analysis part, and forms the hits section in a portion which is a broadband.

An information adjunct which adds the above-mentioned additional information using the formed above-mentioned hits section.

[Claim 15]The information processor according to claim 1 adding the above-mentioned information adding means to the above-mentioned audio signal by making the above-mentioned additional information into a spread spectrum signal.

[Claim 16]An information processing method comprising:

A process of generating an audio signal.

A process of generating additional information.

A process of adding the above-mentioned additional information to the above-mentioned audio signal in a mode which does not influence listening comprehension of a sound by an audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne., and generating an output sound signal.

A process of changing the above-mentioned output sound signal into a sound, and outputting it.

[Claim 17]An information processor comprising:

A voice input means which inputs a sound and acquires an input voice signal corresponding to the sound.

An additional information extracting means which extracts additional information added to the above-mentioned input voice signal.

An information processing means which carries out processing which uses the extracted above-mentioned additional information.

[Claim 18]The information processor according to claim 17, wherein the above-mentioned additional information is information relevant to the above-mentioned sound.

[Claim 19]The information processor according to claim 18 which the above-mentioned sound expresses predetermined natural language, and is characterized by the above-mentioned additional information being information which shows words and phrases contained in the above-mentioned natural language.

[Claim 20]The information processor according to claim 19, wherein the above-mentioned information processing means displays the above-mentioned words and phrases on an indicator using the above-mentioned additional information.

[Claim 21]The information processor according to claim 18 which the above-mentioned sound expresses predetermined natural language, and is characterized by the above-mentioned additional information being information required for recognition of the above-mentioned natural language.

[Claim 22]The information processor according to claim 21, wherein the above-mentioned additional information is a keyword relevant to the above-mentioned natural language.

[Claim 23]The information processor according to claim 21, wherein the above-mentioned additional information is information which shows an intention or feeling expressed by the above-mentioned natural language.

[Claim 24]The information processor according to claim 21, wherein the above-mentioned information processing means performs speech recognition processing which recognizes the above-mentioned natural language from the above-mentioned input voice signal.

[Claim 25]The information processing measure according to claim 24, wherein the above-mentioned additional information is information which specifies a dictionary required for the above-mentioned speech recognition.

[Claim 26]The information processor according to claim 24, wherein the above-mentioned additional information is information for identifying an object domain where the above-mentioned natural language is included.

[Claim 27]The information processor according to claim 18 which the above-mentioned sound expresses a place with natural language, and is characterized by the above-mentioned additional information being position information which shows the above-mentioned place.

[Claim 28]The information processor according to claim 27 using the above-mentioned additional information for the above-mentioned information processing means, and displaying the above-mentioned place on an indicator.

[Claim 29]The information processor according to claim 18 which the above-mentioned sound expresses predetermined natural language, and is characterized by the above-mentioned additional information being information on foreign country language corresponding to the above-mentioned natural language.

[Claim 30]The information processor according to claim 29 using the above-mentioned additional information for the above-mentioned information processing means, and displaying the above-mentioned foreign country language on an indicator.

[Claim 31]The information processor according to claim 18, wherein the above-mentioned sound is voice of human being or an animal and the above-mentioned additional information is information which shows an intention or feeling expressed by the above-mentioned sound.

[Claim 32]The information processor according to claim 18, wherein the above-mentioned sound is a cry of an animal and the above-mentioned additional information is information which shows a kind and a name of the above-mentioned animal.

[Claim 33]The information processor according to claim 32 using the above-mentioned additional information for the above-mentioned information processing means, and displaying a kind and a name of the above-mentioned animal on an indicator.

[Claim 34]The information processor according to claim 17, wherein the above-mentioned additional information is information for identifying a device which outputted the above-mentioned sound.

[Claim 35]The information processor according to claim 17, wherein the above-mentioned additional information is information for identifying a device which has the right to output the above-mentioned sound next.

[Claim 36]The information processor according to claim 17, wherein the above-mentioned information processing means is a control means which controls operation of a robot which imitated a living thing.

[Claim 37]An information processing method comprising:

A process of changing a sound inputted and acquiring an input voice signal.

A process of extracting additional information added to the above-mentioned input voice signal.

A process of performing processing which uses the extracted above-mentioned additional information.

[Claim 38]An information processor comprising:

An audio signal generating means which generates an audio signal.

An additional information generating means which generates additional information.

An information adding means which adds the above-mentioned additional information to the above-mentioned audio signal in a mode which does not influence listening comprehension of a sound by the audio signal, and acquires an output sound signal.

A voice output means which outputs a sound by the above-mentioned output sound signal, a voice input means which inputs a sound and acquires an input voice signal corresponding to the sound, an additional information extracting means which extracts additional information added to

the above-mentioned input voice signal, and an information processing means which carries out processing which uses the extracted above-mentioned additional information.

[Claim 39]The information processor according to claim 38, wherein the above-mentioned information processing means is a control means which controls operation of a robot which imitated a living thing.

[Claim 40]When the above-mentioned input voice signal including information which solves the above-mentioned inquiry from the above-mentioned voice input means after outputting a sound by the above-mentioned output sound signal concerning an inquiry from the above-mentioned voice output means is not acquired, An information storing means which stores information on the above-mentioned inquiry as information on an unsolved inquiry, Based on information on an inquiry of above-mentioned un-solving [which is stored in the above-mentioned information storing means], The information processor according to claim 38 having further a re output control means to which a sound by the above-mentioned output sound signal which starts an inquiry of above-mentioned un-solving from the above-mentioned voice output means to arbitrary timing is made to output.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention]This invention relates to the information processor and information processing method which enable communication of information with a sound. By adding additional information to an audio signal in detail in the mode which does not influence listening comprehension of the sound by the audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne., generating an output sound signal, changing the output sound signal into a sound, and outputting it, Information presentation which is easy to understand with a sound can be given to a user, and the information processor etc. which enabled it to realize the exchange of additional information or the information that accuracy with few errors is high are started between devices.

[0002]

[Description of the Prior Art]There is the method of connecting between devices with a cable etc. or communicating on radio as a correspondence procedure between two or more information processors, using infrared rays, an electric wave, etc. Generally, although two devices are connected by the couple 1, communication between three or more devices can also be performed by using a device with a special hub etc. on the way.

[0003]In the input from a user to an information processor, signal transduction through natural language can be performed by using speech recognition technology. It is also possible by conducting syntax analysis and a semantic analysis for the natural language by a receiver to provide various services for a user.

[0004]As an output to a wide range user, the voice response through a loudspeaker is used widely. The device for which a sound's spreading broadly, a loudspeaker, etc. are needed is carried out by the system of the yard announcement of a station, inside of a shop, etc., etc. from the ability to realize comparatively cheaply, for example.

[0005]

[Problem(s) to be Solved by the Invention]However, in the method of connecting between devices by cables, such as a cable, the terminal for connecting a cable to a device is needed, and the cable for connection is needed. In order to communicate among three or more devices, in addition, a device with a special hub etc. is needed. Thus, prior preparation was needed for communicating and the technical problem that it could not communicate immediately occurred.

[0006]In the method of using infrared rays, an electric wave, etc., it needed to have the special mechanism, only in order to communicate, and the technical problem that it became the hindrance of the miniaturization of a device or low-pricing occurred.

[0007]Generally a user does not know what kind of data is exchanged between devices. In order for the user to have understood this, a screen display, voice response, etc. had the technical problem that a means different from a channel had to be used.

[0008]When voice input is used as an input interface to an information processor, with the present art, a user's utterance cannot be recognized correctly thoroughly. Since the number of recognized vocabularies had big influence on recognition performance especially, if possible, the number of vocabularies needed to be restricted to performing high-precision recognition, and it needed to devise narrowing down the object of recognition etc. In the present speech recognition, information, including the feeling etc. which are inherent in a user's utterance, can

hardly be acquired. Also in syntax analysis or a semantic analysis, it is dramatically difficult to draw what reflected a speaking person's input intention thoroughly.

[0009]In the information processor provided with the voice input/output, although it is possible to exchange information interactively and information exchange between devices can be performed by exchanging an inquiry mutually, when I do not understand as a result of an inquiry, no information exchange is made. For this reason, in order to have asked again, the user needed to do voice input again, the recognition error etc. might arise and there was a problem of becoming a burden for a user there.

[0010]When voice response is used as an information presenting means to a user, in the present announcement, information does not get across to those who do not hear an ear at all, for example. Since it is furthermore expressed with natural language, information is not transmitted also when there is a difference in the language which a user uses. The place etc. which the user was not well versed in the geography of the place in the inside-of-a-shop announcement with a sound in many cases, and were expressed by means of language are unclear.

[0011]So, it aims at providing the information processor and information processing method which can solve the technical problem mentioned above in this invention.

[0012]

[Means for Solving the Problem]An audio signal generating means in which an information processor concerning an invention of claim 1 generates an audio signal, It has an additional information generating means which generates additional information, an information adding means which adds additional information to an audio signal in a mode which does not influence listening comprehension of a sound by an audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne., and generates an output sound signal, and a voice output means which outputs a sound by an output sound signal.

[0013]An information processor concerning an invention of claim 17 is provided with a voice input means which inputs a sound and acquires an input voice signal corresponding to the sound, an additional information extracting means which extracts additional information added to an input voice signal, and an information processing means which carries out processing which uses extracted additional information.

[0014]An audio signal generating means in which an information processor concerning an invention of claim 38 generates an audio signal, An additional information generating means which generates additional information, and an information adding means which adds additional information to an audio signal in a mode which does not influence listening comprehension of a sound by an audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne., and acquires an output sound signal, A voice output means which outputs a sound by an output sound signal, and a voice input means which inputs a sound and acquires an input voice signal corresponding to the sound, It has an additional information extracting means which extracts additional information added to an input voice signal, and an information processing means which carries out processing which uses extracted additional information.

[0015]Communication with a sound is attained in this invention. A sound in this case is based on an output sound signal acquired by adding additional information to an original audio signal in a mode which does not influence listening comprehension of a sound by an audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne. Additional information is information relevant to an original audio signal, for example. For example, an attack portion in an audio signal is removed, and the hits section is formed in a portion which is a broadband, and addition of additional information is performed using the hits section. For example, additional information is added to an audio signal as a spread spectrum signal. Information presentation which is easy to understand with a sound can be given to a user by this, and it becomes possible between devices to realize an exchange of additional information or information that accuracy with few errors is high.

[0016]

[Embodiment of the Invention]Hereafter, this embodiment of the invention is described, referring to drawings. Drawing 1 shows a general view of the information processor 10 as a 1st embodiment. The loudspeaker 12 for outputting a sound is formed, and the microphone 13 for inputting a sound is formed in the main part 11 of this information processor 10. The talk switch

14 operated when inputting a sound via the microphone 13 is formed in this main part 11. In this case, when the talk switch 14 is operated by the user, the voice input from the microphone 13 becomes possible.

[0017]In order to display GUI (Graphical User Interface) of a program on the center section, the indicator 15 is formed in the main part 11. On the surface of this indicator 15, when a user contacts using the touch pen 1 or a finger, what is called the touch panel (touch tablet) 16 that outputs the signal corresponding to the directed position is arranged.

[0018]Here, the touch panel 16 is constituted by transparent materials, such as glass or resin. Therefore, the user can see the picture displayed on the indicator 15 through the touch panel 16. The user can input a predetermined character into the touch panel 16 using the touch pen 17, or can perform selection or execution of a predetermined object (icon) currently displayed on the indicator 15.

[0019]Drawing 2 shows the circuitry of the information processor 10. The internal bus 20, CPU (central processing unit) 21, ROM(read only memory) 22, RAM(random access memory) 23, the display control part 24, the input interface 25, and the speech synthesis section 26 are connected mutually. Thereby, each part can deliver and receive data via the internal bus 20. CPU21 is made as [perform / various kinds of processings] according to the program or various kinds of data which are memorized by ROM22 or RAM23.

[0020]The display control part 24 generates the data of the picture displayed on the indicator 15 corresponding to the information supplied from CPU21, and displays the picture on the indicator 15. The input primary detecting element 27 is made as [supply / detect the input of the touch panel 16 or the talk switch 14, and / to the input interface 25 / a corresponding manipulate signal]. The A/D conversion part 28 changes into a digital signal the audio signal outputted from the microphone 13 from an analog signal, and is made as [supply / the input interface 25].

[0021]The input interface 25 receives the audio signal supplied from the A/D conversion part 28, or the manipulate signal supplied by the input primary detecting element 27, and is made as [supply / CPU21]. If an audio signal is inputted via the microphone 13, the A/D conversion part 28, and the input interface 25, CPU21 will perform extracting processing of the additional information added to the audio signal with reference to the data memorized by ROM22 or RAM23.

[0022]The speech synthesis section 26 generates synthesized speech based on a parameter and text data required for the voice synthesis supplied from CPU21, and is made to output it via the loudspeaker 12. The speech synthesis section 26 is used also when playing the sound recorded by RAM23 via the microphone 13. When the additional information which should be transmitted to other devices from CPU21 is supplied, the speech synthesis section 26 performs attached processing of the additional information to synthesized speech or the recorded sound, and is made to output it via the loudspeaker 12.

[0023]Drawing 3 shows the functional block diagram of the interior action of CPU21 mentioned above. The information processing section 30 performs various information processing, and performs processing as an application program in which individual schedule management is performed, in this embodiment.

[0024]The information adjunct 31 adds the specific additional information IFa provided from the information processing section 30 to the audio signal SAa provided from the information processing section 30, and generates the output sound signal SAout. Addition of here as opposed to the audio signal SAa of the additional information IFa is performed in the mode which does not influence listening comprehension of the sound by the audio signal SAa.

[0025]For example, attached processing is performed using the method indicated to JP,10-162501,A. In this case, the following attached processing is performed in the information adjunct 31. First, a spectrum analysis is conducted about the section of the rapid length of the audio signal SAa as which rose, the large amplitude portion was detected as an attack, and the audio signal SAa was determined beforehand. And the attack portion in the audio signal SAa is removed, and the hits section is formed in the portion which is a broadband, the additional information IFa is added using the interception section, and the output sound signal SAout is generated. The influence of an attack portion which it has on tone quality is great. The click can become difficult to be heard, so that bandwidth is wide. Therefore, the click by hits can be made

hardly audible by carrying out attached processing as mentioned above, and it becomes possible for there to be no degradation of tone quality and to add the additional information IFa to the audio signal SAa.

[0026]For example, when a certain big audio signal in human being's aural characteristic exists, or the signal of the low near the frequency cannot be heard, it may be made to perform attached processing using the "auditory masking characteristic" of being very hard to be audible. It may be made to superimpose additional information on an audio signal by the spread spectrum which furthermore attracts attention these days. In this case, it adds to the audio signal SAa by making additional information IFa into a spread spectrum signal.

[0027]The information extraction part 32 carries out processing which extracts the additional information IFb from the input voice signal SAin (it corresponds to the output sound signal SAout mentioned above). It depends for the processing performed by this information extraction part 32 on the disposal method made in the information adjunct 31. Here, additional information may not be inserted.

[0028]The information processing section 30 is a portion which performs a actual application process, and is made as [supply / supply the audio signal SAa and the additional information IFa to the information adjunct 31, and / from the information extraction part 32 / the additional information IFb or the audio signal SAb].

[0029]Next, the flow of the processing performed with the functional block diagram shown in drawing 3 is explained by making information processing aiming at schedule management into an example. Drawing 4 and drawing 5 show the flow of the processing, and when it sends out a sound (information) outside, respectively, they express the case where a sound (information) is inputted from the exterior. Although the processing which performs individual schedule management is used in the following explanation, also in other examples of application, it can carry out in the same procedure.

[0030]Voice response operation is explained with reference to drawing 4. First, in Step S40, a user chooses the information which should be transmitted using input interfaces, such as the indicator 15, the touch panel 16, and the touch pen 17. For example, the object "5/5 15:00 meeting" showing one item of the schedule of the indicator 15 is chosen. And a natural text [in the information processing section 30] in Step S41 which expresses an item with the selected user (utterance sentence), for example, "it is scheduled to hold a conference from 15:00 on May 5", is generated, and the additional information of identifiable data representation is further generated inside a system.

[0031]In continuing Step S42, the synthesized speech signal SAa for reading out the utterance sentence generated at Step S41 is generated by the speech synthesis section 26. In Step S43, the additional information IFa generated at Step S41 is added to the synthesized speech signal SAa generated by the speech synthesis section 26 by the information adjunct 31. Under the present circumstances, attached processing is performed in the mode which does not influence listening comprehension of the sound by the synthesized speech signal SAa, as mentioned above. That is, attached processing is performed by the technique [change of the audio signal SAa by addition of the additional information IFa is not discriminable by human being's acoustic sense, or] of being difficult to identify.

[0032]Finally, in Step S44, the output sound signal SAout including the additional information IFa generated at Step S43 is supplied to the loudspeaker 12, and the sound by the output sound signal SAout is outputted from this loudspeaker 12. In this way, the sound outputted can understand the utterance sentence which human being who hears it mentioned above with natural language. In other information processors 10, it becomes more possible than the input voice signal SAin concerning the sound to understand the utterance sentence mentioned above by extracting the additional information IFb (it is the same as IFa).

[0033]Next, voice input operation is explained with reference to drawing 5. First, in Step S50, alien-frequencies voice is taken in with the microphone 13, and the input voice signal SAin is acquired. Introduction of alien-frequencies voice considers as a start signal that the talk switch 14 is pushed by the user, or takes in while being pushed, and considers it as the section, or detects the sound more than a certain level, and performs the processing.

[0034]Continuing extracting processing of the additional information IFb included in the input

voice signal SAin by the information extraction part 32 in Step S51 is performed. As a result of processing, when the additional information IFb is not extracted, processing is ended. When the additional information IFb is extracted, it progresses to Step S52 and the processing based on the additional information IFb is made in the information processing section 30. For example, when a sound which was explained by drawing 4 is taken in with the microphone 13, to the input voice signal SAin. The data expressing "5/5 15:00 meeting" is inserted as the additional information IFb, by extracting this additional information IFb, the processing for the addition/renewal of information has done enough, and processing which asks the renewal permission of data to a user is made.

[0035]Next, with reference to drawing 6, the two-way communication operation between the 1st information processor 10 (device A) and the 2nd information processor 10 (device B) is explained. The device A is a device of the side which asks about a schedule, and the device B is a device of the side which answers the inquiry.

[0036]First, in the device A, the user Ua inputs a command which asks "what time tomorrow's meeting is." The input of this command is performed using the indicator 15, the touch panel 16, and the touch pen 17. When the device A has a speech recognition function, it may be made to input an above-mentioned command with a sound from the microphone 13. In Step S60, with the device A according to the user's Ua alter operation in the same procedure as the operation shown in the example of above-mentioned drawing 4. The output sound signal SAout included the additional information IFa "what time tomorrow's meeting is" is generated, and the sound by the output sound signal SAout is turned to the device B side, and is outputted. By hearing the audio signal, the user Ua of the device A and the user Ub of the device B can understand clearly the processing which the device A tries to perform.

[0037]In the device B, in Step S61, the sound outputted from the device A is taken in from the microphone 13, and it processes by the same operation as the example shown by drawing 5. The command which asks "what time tomorrow's meeting is" is inserted in the input voice signal SAin as the additional information IFb, and the additional information IFb of this command is extracted in Step S62 in the device B. And in the device B, in continuing Step S63, information retrieval to the time of tomorrow's meeting is performed, and a saying "from 15:00" result is obtained as a result of an inquiry the information processing based on the additional information IFb of the extracted command, and here.

[0038]The result of the inquiry obtained at Step S63 is made as [choose / as information in Step S40 of drawing 4 which should be transmitted / as it is / automatically]. Or the device B asks for the user's Ub permission, and it may be made as [choose / only when it is allowed]. In the device B, if the information which should be transmitted is decided, generation of the output sound signal SAout will be performed in continuing Step S64 by the same operation as drawing 4. In S64, drawing 6 shows as an example the case where the response sentence that it is from ** "" is not understood" 15 o'clock" is generated. [**] Here, the output sound signal SAout included not only a mere audio signal but the additional information IFa expressing a clear schedule is generated, and in Step S65, the sound by the output sound signal SAout is turned to the device A side, and is outputted.

[0039]In the device A, the additional information IFb expressing the schedule which takes in the sound outputted from the device B from the microphone 13, and is continuously included in the input voice signal SAin in Step 67 is extracted in Step S66. And when the additional information IFb is the information ** "from 15:00", it progresses to Step S68 and addition/ updating to the schedule information of the device B are automatically performed on the basis of permission of the user Ua. Processing which incorporates into the device A the information stored in the device B from the first by interactive exchange of the above devices A and the device B according to the question from the device A is realizable.

[0040]the extracted additional information -- **, when it is the information "it does not understand", it progresses to Step S69 and information required in order to generate the output sound signal SAout of the inquiry which the device A generated with the information that an inquiry is unsolved, or it is stored in RAM23 in the device A. The user Ua can see via the indicator 15, and the stored information is made as [carry out / sound / by the output sound signal SAout / by choosing the information via the touch panel 16 / a re output]. Thus, the

inquiry which was not solved in the dialog between devices is stored in the device, and it becomes possible to reduce a user's work because it can be made to carry out the re output of the sound which starts the inquiry at the arbitrary times of after that.

[0041]Although drawing 6 explained two-way communication operation of 1 to 1, since the sound is used as an information transmission medium, one-pair Oshi can also be communicated simultaneously. It enables this to transmit simultaneously the schedule information of "5/5 15:00 meeting" emitted from the one information processor 10 to two or more information processors 10 like the example raised by previous drawing 4 and 5.

[0042]It may be made for the identifier of the information processor 10 which outputs the sound by the audio signal as additional information added to an audio signal to be included. Thereby, the information processor 10 which took in the sound from the microphone 13 can know which is the information processor 10 which outputted the sound, and can perform processing by the information processing section 30 according to this. For example, in an above-mentioned example, if it is the schedule update information from the reliable information processor 10, information will be updated automatically, otherwise, the permission check to a user will be asked. Also in communication between the devices of above-mentioned plurality, each device can know the sound from which device it is.

[0043]Interactive data communications which were raised with drawing 6 as an example are also possible. Since a possibility that two or more devices will output a sound simultaneously arises when performing such a dialog between plural equipments, it is made for the identifier of a device which gives the right to permit the next utterance as additional information to be included. If the device which incorporated the sound differs in the identifier contained in the input voice signal SAin from its identifier, it will be made as [carry out / voice response], and it can prevent two or more devices carrying out voice response simultaneously using this additional information.

[0044]It is also possible to exchange a memo with a sound among two or more information processors 10 which applied this invention. In this case, by adding some keywords to which the contents of the memo belong to the audio signal showing a memo as additional information, the device of a receiver can know the field where the memo expressed by that sound belongs, and can carry out classification arrangement of that memo automatically by a receiver.

[0045]According to this invention, it is possible to perform long-distance or wide range communication through voice media. In this invention, since the sound is used for information transmission, it becomes possible via the media using a sound, for example, television, radio, etc., to perform a wide range offer of information. For example, the information on the product can be acquired from CM inserted into television or radio.

[0046]Next, a 2nd embodiment of this invention is described.

[0047]Although the information processor 10 as a 1st embodiment mentioned above is not provided with the speech recognition function, the information processor 10 as this 2nd embodiment is provided with the speech recognition function. The circuitry of the information processor 10 in this 2nd embodiment is the same as that of the information processor 10 in a 1st embodiment (refer to drawing 2). If an audio signal is inputted via the microphone 13, the A/D conversion part 28, and the input interface 25, After processing which extracts the additional information IFb from the input voice signal SAin like a 1st embodiment is performed, CPU21, With reference to voice learned data and dictionary information which are memorized by ROM22 or RAM23, it is made as [perform / to the audio signal SAb / speech recognition processing].

[0048]In the present speech recognition technology, when recognition rates are not 100% but many, an error will be included. It is dramatically difficult to extract the delicate intention included in the sound, and feeling by the system side. For example, two texts ["be / tomorrow / from 5:00 / it / a meeting"] "whether it will be a meeting from 5:00 tomorrow" -- an end -- it is -- although the meanings differ greatly by whether to buy and there to be in being attached, in speech recognition, it is only a difference for only 1 sound, and is easy to produce a recognition error.

[0049]Then, in the information processor 10 with a speech recognition function, the additional information IFa for distinguishing the intention of a "inquiry", a "question", etc., etc. is added to the audio signal SAa. By adding the information which shows the intention of an "inquiry" as the

additional information IFa from the device A to the audio signal SAa to the device B, if it explains with reference to above-mentioned drawing 6 used in order to explain the two-way communication operation between devices, the "information processor 10 of a receiver — an end — it is " — even if it misses at the time of speech recognition, it is not interpreted as it being the intention of a "decision", and interpretation/processing can be performed under the condition that it is a sentence of an inquiry.

[0050]When using at the time of the information inputting from a user to the information processor 10, between devices, the synthesized speech generated within the thing which recorded the user's sound, or the device is used using a user's sound. Since these audio signals differ in the audio characteristic, in order to perform accurate speech recognition, respectively different voice learned data is needed for ROM22 or RAM23, and a big storage capacity is needed. Then, in communication between human being and the information processor 10, using speech recognition in communication between devices, give all information required for processing by additional information, or, The information etc. which specify a dictionary required for the keyword or recognition showing the object domain where the additional information for making it avoid or reduce, for example, an intention and a text, is included [that a recognition error arises and] are given.

[0051]For example, the voice response from an information processor is related with a "travel", and moreover, when meaning a "inquiry", the intention information of the keyword and a "inquiry" a "travel" is inserted in an audio signal as additional information. In the information processor 10 which takes in the sound by the audio signal, what extracted the "travel" and the "inquiry" as additional information first, then was suitable for the "travel" and the "inquiry" as dictionary information for speech recognition is chosen, and speech recognition is performed. This will restrict an object domain and speech recognition can be performed with more sufficient accuracy.

[0052]Next, a 3rd embodiment of this invention is described.

[0053]Although this invention was applied to the personal information management apparatus in 1st and 2nd embodiments mentioned above, this invention can also be applied to a robot.

[0054]The robot is equipped with the speech recognition mechanism and the voice synthesis mechanism, and it can have a dialog using natural language mutually. In the utterance to a robot from human being, it is made as [aim at / by using speech recognition technology / mutual intention understanding]. However, the dialog of robots is made as [exchange / actually / information / based on the additional information added to the audio signal], although it seems to be seemingly made by natural language. Thereby, between robots, signal transduction with high reliability which avoided the error of speech recognition can be performed, and human being who is observing it can also catch the contents as synthesized speech, and communication with higher compatibility and reliability can be realized.

[0055]Also in the industrial world, a team is composed by two or more robots, and an event called RoboCup (RoboCup) of playing a game in soccer by 2 teams is held every year. Here, although the game of soccer is held by the robot of various gestalten, the spectator who observes it does not know whether a robot considers what, exchanges what kind of information, and is moving. Although various information exchange is made by robots, it cannot be known from the spectator side.

[0056]A general view of the game of the soccer by the robot which applied this invention is shown in drawing 7. In order to perform soccer, the field 70 is equipped with the goal 71, and the robots 73, 74, and 75 are playing involving the ball 72. Drawing 7 shows a part of field 70, and the robot and the goal which do not appear in drawing 7 exist.

[0057]The robots 73, 74, and 75 are equipped with the voice input/output part, and a robot and others considers what and it is made as [know / what kind of information exchange is carried out / the spectator 76] because it is made to exchange information by uttering a sound mutually. At drawing 7, the utterance "run before the goal" is made by the robot 74 from the robot 73, and the spectator 76 can know by hearing it as a sound. The additional information IFa is added to the output sound signal SAout for outputting the sound which the robot 73 utters, and signal transduction from the robot 73 to the robot 74 can be performed in high accuracy.

[0058]For example, to the output sound signal SAout for outputting the sound "run before the

goal" to the robot 74 from the robot 73 as the additional information IFa. It is made to perform required signal transduction between robots by adding the identifier showing the robot 73 which spoke, the identifier showing the robot 74 made into the object of a command, its concrete command, etc. It is also possible to exchange the mutual position information between robots by giving the position coordinate as the additional information IFa corresponding to the sound "goal before." Although a sound has a certain amount of directivity, it has the character spread in the wide range. Therefore, it is possible to take in a sound also in the robot 75 in drawing 7, and to extract the additional information. However, in the robot which has not appeared in drawing 7 and which separated further, a sound cannot fully be taken in and information exchange with the sound from the robot 73 cannot be performed. This is the same character as the game by actual human being, and can realize the game nearer to reality using a robot.

[0059]When thinking that it is more desirable not to utter natural language like an animal type robot, a cry etc. may be used as a sound. In this case, it is dramatically difficult to recognize a natural cry in the present speech recognition technology, and to distinguish those contents. Although human being does not understand as a language clear only by judging from a cry by adding the information which shows an intention and feeling of the cry as the additional information IFa to the output sound signal SAout for outputting a cry, clearer intention understanding can be aimed at between robots. Thereby, although a certain language is exchanged between animals, the robot near the relation of actual human being and the animal that human being does not understand in detail is realizable.

[0060]A virtual zoo is realizable by applying this invention. It is not a actual animal and is constituted from this virtual zoo by the cry of an animal, an image, or the robot. Various kinds of animals are stored as information here, and the user can see the animal which cannot be seen usually.

[0061]Although the human being can distinguish the kind of animal from the cry of an animal to some extent, in the present speech recognition, it is mainly taken into consideration hardly to sounds, such as a cry, for recognition of human being's language. Therefore, it is difficult to discriminate a kind from a cry. A user does not understand the kind or name for the animal which does not not much have familiarity, either.

[0062]Then, information, including the kind of the animal, a name, etc., is added to the output sound signal SAout which outputs a cry as the additional information IFa. The user can have the information processor 10 in this invention, and, thereby, can get that kind and name of that that were added as additional information from the cry of each animal. With an information processor, more detailed information can be shown to a user based on the kind and name. Thereby, a virtual zoo which shows a user the kind, the name, and also the more detailed information on the animal is realizable by using the cry of an animal as a key.

[0063]Although **** described the case where only additional information was used for the signal transduction between robots, of course, speech recognition may be performed and you may be a gestalt using additional information as the assistance.

[0064]Next, a 4th embodiment of this invention is described.

[0065]As an example of a system using the information processor mentioned above, the announcement system used by the yard announcement of a station, inside of a shop, etc. can be considered. A general view of the announcement system is shown in drawing 8. The server 80 is for performing a yard announcement, and is made as [connect / with the loudspeaker 81 etc. which were installed in yard every place]. The user 82 is carrying the information processor 83 used as a client side, hears the sound outputted from the loudspeaker 81 with his ear, and he takes in the sound with the information processor 83.

[0066]The server 80 is provided with the function by the side of the voice response of the information processor 10 shown in drawing 2, and the output interface portion is connected with two or more loudspeakers 81. The server 80 has the function to carry out [sound] various information, including merchandise information, such as the time of arrival of a missing child's information and the next train, and a bargain sale, etc., with which it is necessary to provide the large area of premises, and to transmit it to the user 82 via the loudspeaker 81 for example. On the other hand, the information processor 83 has a general view which is provided with the function by the side of the voice input of the information processor 10 shown in drawing 2, for

example, was shown in drawing 1.

[0067]The operator is operating the server 80, and when the phenomenon which should be announced occurs at any time, it performs treating operation which suited it. The sound by the natural language with which the server 80 transmits and the user 82 of premises is provided through the loudspeaker 81 is synthesized speech created [utterance, the sound currently recorded beforehand, or if needed] for an operator. However, neither the announcement of the timetable in a station yard nor the announcement of opening closing of a store needs an operator, but, for example, may be made to be outputted about what may carry out voice response automatically according to the schedule defined beforehand automatically.

[0068]As additional information added to the audio signal concerning these sounds, information which expresses the contents which the natural language means, for example can be used.

Although it is dramatically difficult to carry out speech recognition of the sound of an announcement itself in the information processor 83, the contents of an announcement can be checked by adding the contents to mean as additional information using the indicator 15 of the information processor 83. Or an ear cannot be heard, those who cannot be heard easily also have the effect that announcing simultaneously is possible.

[0069]In an announcement like "... come before the third floor elevator" which is made by a missing child's information, there is a problem of spending time in finding the place. As additional information to such an announcement, the position information "in front of the third floor elevator" can be provided in this case. It is possible to perform more concrete information presentation to the user 82 by displaying the map information based on the position information on the indicator 15 in the information processor 83 which received it.

[0070]The information which expresses languages, such as English and French, to the announcement of a different language from the natural language which expresses an announcement as additional information, for example, Japanese, is given. Thereby, the same information can be transmitted with a sound also to the foreigner who does not understand the language of the country. The user 82 who cannot understand the meaning of the sound itself announced on premises from the difference in language, By using the information processor 83, if the information expressing translation to a native language is added to the audio signal, you can understand it as a native language via the indicator 15 or the loudspeaker 12.

[0071]In the above-mentioned embodiment, as for this invention, although this invention was applied to the portable information processor 10, the robots 73-75, or the server 80 as shown in drawing 1, it is needless to say that it is applicable like other devices.

[0072]

[Effect of the Invention]According to this invention, add additional information to an audio signal in the mode which does not influence listening comprehension of the sound by the audio signal of opposite *Perilla frutescens* (L.) Britton var. *crispa* (Thunb.) Decne., and an output sound signal is generated, The output sound signal can be changed into a sound, and can be outputted, information presentation which is easy to understand with a sound can be given to a user, and the exchange of additional information or the information that accuracy with few errors is high can be realized between devices. Since the sound is used as communication media, information can be simultaneously exchanged not between communication of the existing directive couple 1 but between the devices of nearby plurality. Long-distance communication and communication for the large-scale number can be performed by using voice media, such as television, radio, and a telephone, as a channel.

[Translation done.]